The EMBO Journal (2008) 27, 1172–1181 | © 2008 European Molecular Biology Organization | All Rights Reserved 0261-4189/08
www.embojournal.org

THE EMBO JOURNAL

Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD

Marcin Nowotny[1,3], Susana M Cerritelli[2], Rodolfo Ghirlando[1], Sergei A Gaidamakov[2], Robert J Crouch[2] and Wei Yang[1,*]

[1]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA and [2]Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

Human RNase H1 contains an N-terminal domain known as dsRHbd for binding both dsRNA and RNA/DNA hybrid. We find that dsRHbd binds preferentially to RNA/DNA hybrids by over 25-fold and rename it as hybrid binding domain (HBD). The crystal structure of HBD complexed with a 12 bp RNA/DNA hybrid reveals that the RNA strand is recognized by a protein loop, which forms hydrogen bonds with the 2′-OH groups. The DNA interface is highly specific and contains polar residues that interact with the phosphate groups and an aromatic patch that appears selective for binding deoxyriboses. HBD is unique relative to non-sequence-specific dsDNA- and dsRNA-binding domains because it does not use positive dipoles of α-helices for nucleic acid binding. Characterization of full-length enzymes with defective HBDs indicates that this domain dramatically enhances both the specific activity and processivity of RNase H1. Similar activity enhancement by small substrate-binding domains linked to the catalytic domain likely occurs in other nucleic acid enzymes.

The EMBO Journal (2008) **27,** 1172–1181. doi:10.1038/emboj.2008.44; Published online 13 March 2008
Subject Categories: structural biology
Keywords: dsRNA; processivity; RNA/DNA hybrid; RNase H; specificity

Introduction

RNases H are endonucleases that specifically hydrolyse the RNA strand in an RNA/DNA hybrid with limited sequence preference. RNA/DNA hybrids exist transiently during DNA replication and transcription as well as reverse transcription in certain viruses. Hybrids of RNA and DNA can also form between mRNAs and artificially introduced complementary oligo-deoxyribonucleotides, which lead to mRNA degradation by cellular RNases H and reduced gene expression. Two types of RNase H, H1 and H2, are distinguished by biochemical properties and substrate preference (Ohtani *et al*, 1999). RNase H1 is present in retroviruses and species ranging from bacteria to humans. In unicellular organisms, deletion of RNase H1 merely reduces cell growth rate probably because RNAs in RNA/DNA hybrids can be removed by other nucleases (Crouch and Cerritelli, 1998; Kogoma and Foster, 1998). However, RNase H1 knockout mice die during embryogenesis due to defective mitochondrial DNA replication (Cerritelli *et al*, 2003). All eukaryotic and some bacterial RNases H1 contain one or two copies of an ~50 residue N-terminal dsRNA and RNA/DNA hybrid binding domain (dsRHbd), which is tethered to the C-terminal catalytic domain (RNase HC) by a linker of variable length (Cerritelli and Crouch, 1995).

The catalytic domain of RNase H1 consists of ~150 residues including four absolutely conserved carboxylates in the active site. Catalysis occurs through a two-metal ion mechanism (Nowotny *et al*, 2005; Nowotny and Yang, 2006). Structures of the catalytic domains of human and *Bacillus halodurans* RNases H1 (RNases HC) in complex with an RNA/DNA substrate have been determined (Nowotny *et al*, 2005, 2007) and reveal two modes of RNA/DNA hybrid recognition. The RNA strand is always recognized through the 2′-OH groups; the DNA strand is recognized on the basis of either the B-form conformation or the absence of 2′-OH groups. Both bacterial and human RNases HC have a conserved phosphate-binding pocket, which accommodates a phosphate group from the non-cleaved strand. This interaction requires a narrow minor groove and thus the B-form conformation, which is adopted by DNA but not RNA. In human RNase HC, there is an additional substrate-binding element called the basic protrusion. It contains a DNA-binding channel, which binds A-form nucleic acids and excludes RNA by close packing of the indole ring of a tryptophan against a deoxyribose of DNA (Nowotny *et al*, 2007).

The small N-terminal domain has a highly conserved amino-acid sequence and is markedly basic with a pI of ~10.0 (Cerritelli and Crouch, 1995; Cerritelli *et al*, 1998). Initial binding studies showed that this domain bound both dsRNA and RNA/DNA hybrids, and it was named dsRHbd (Cerritelli *et al*, 1998). Although the catalytic activity resides in the C-terminal domain, dsRHbd enhances the affinity for RNA/DNA hybrids and processivity of mouse and human RNases H1. Surface plasmon resonance (SPR) analyses indicated that two molecules of the full-length RNase H1 could bind to a 12 bp hybrid, whereas the same hybrid only accommodated one RNase HC without the dsRHbd (Gaidamakov *et al*, 2005). The dsRHbd was thus suggested to promote dimerization of eukaryotic RNases H1 and thereby increase the processivity of the enzyme. A Trp and two Lys residues conserved in dsRHbd were identified as critical for both dimerization and processivity of mouse RNase H1 (Gaidamakov *et al*, 2005).

*Corresponding author. Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
Tel.: +1 301 402 4645; Fax: +1 301 496 0201;
E-mail: Wei.Yang@nih.gov
[3]Present address: Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, 4 Trojdena Street, Warsaw, Poland

Received: 15 December 2007; accepted: 19 February 2008; published online: 13 March 2008

The structure of dsRHbd from yeast RNase H1 was determined by NMR and was shown to consist of a short three-stranded anti-parallel β-sheet and two helices (Evans and Bycroft, 1999). The domain is monomeric in solution and its fold is different from known dsRNA- and dsDNA-binding modules. The molecular mechanisms for its binding to both dsRNA and RNA/DNA hybrid and for its role in catalysis remain uncertain. We report here the binding preference of the dsRHbd of human RNase H1, the crystal structure of its complex with an RNA/DNA hybrid and its role in the RNase H1 activity.

Results and discussion

HBD prefers to bind RNA/DNA hybrid

Previous binding studies of dsRHbd were conducted with whole-cell lysates (Cerritelli *et al*, 1998) or in the context of full-length RNases H1 (Cerritelli and Crouch, 1995) and showed no distinction between dsRNA and RNA/DNA hybrids. For biochemical and structural characterization, the N-terminal dsRHbd (residues 27–76) of human RNase H1 was expressed in *Escherichia coli* and purified using affinity and ion-exchange chromatography. Affinities of the isolated domain for a 12 bp RNA/DNA hybrid, dsRNA and dsDNA of the same sequence (with interchanges of T in DNA and U in RNA) (Supplementary Table I) were studied by electrophoretic mobility shift assay (EMSA), gel filtration and sedimentation equilibrium centrifugation. In spite of dissociation of protein–nucleic acid complexes during electrophoresis as indicated by the smearing, EMSA showed that dsRHbd has the highest affinity for RNA/DNA hybrids, lower affinity for dsRNA and the lowest for dsDNA (Figure 1A). In gel filtration experiments, only the RNA/DNA formed detectable complexes with the dsRHbd (data not shown). Quantitative analyses by analytical ultracentrifugation revealed the K_d for the 1:1 complexes of dsRHbd-RNA/DNA to be below

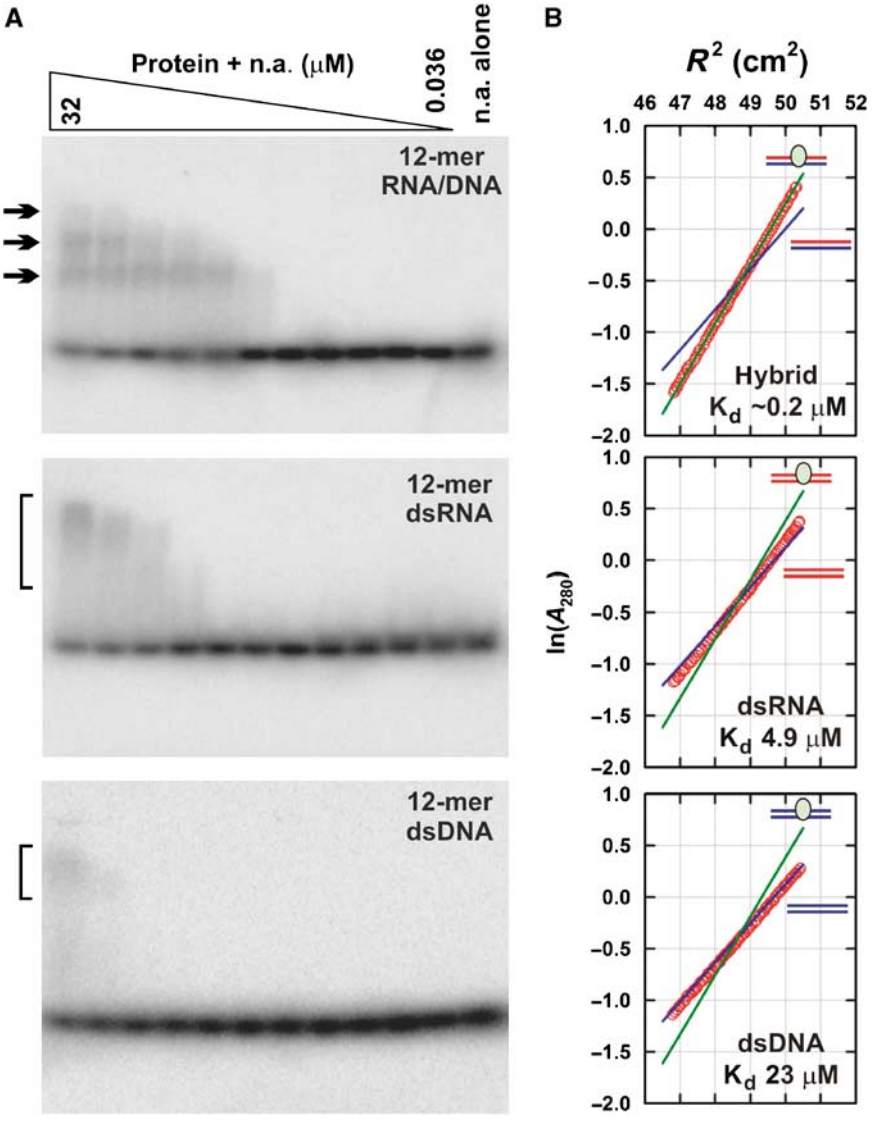

Figure 1 Hs-HBD preferentially binds to RNA/DNA hybrids. (**A**) EMSA of nucleic acid (n.a.) binding. Concentrations from 32 to 0.036 μM of the 1:1 mixtures of Hs-HBD and 12 bp RNA/DNA, dsRNA or DNA of equivalent sequences were resolved on a native TBE polyacrylamide gel. Bands corresponding to HBD–nucleic acid complexes are indicated by arrows or brackets. (**B**) Sedimentation equilibrium profiles of 1:1 mixture of ~5 μM Hs-HBD and the 12 bp RNA/DNA (red/blue), dsRNA (red) or dsDNA (blue) at 22 k.r.p.m. and 4°C. Data (red circles) were collected at 280 nm. The solid blue lines show the slope expected for the free duplex nucleic acid, and the solid green lines the slope expected for the 1:1 protein–nucleic acid complex. Complete data analyses are detailed in the Supplemental data.

0.2 μM (Figure 1B, Supplementary Figures 1, 2, and 3) and the K_d values for the 1:1 complexes of dsRNA and dsDNA to be 4.9 and 23.0 μM, respectively. Therefore, the N-terminal domain of Hs-RNase H1 binds preferentially to RNA/DNA hybrids by at least 25-fold and is renamed as HBD.

Three distinct bands of HBD-RNA/DNA hybrid complexes were observed with increasing concentrations of complexes in EMSA (Figure 1A), indicating binding of more than one HBD to a 12 bp duplex. Interestingly, the fastest moving band among the three, which presumably corresponded to the 1:1 complexes, occurred first and persisted throughout the concentration increase, which suggested that the 1:1 complexes might be tightest. Analytical ultracentrifugation (AUC) analysis corroborated the finding with detection of only 1:1 complexes at various loading concentrations as long as the molar ratio was 1:1 (Supplementary Figure 1). Appearance of two additional slower-migrating bands in EMSA (presumably 2:1 and 3:1 complexes) and free DNA at the bottom of the gel might result from non-equilibrium nature of gel electrophoresis. Two HBDs on a single 12 bp hybrid were detected by size exclusion chromatography (data not shown), and 2:1 complexes of dsRNA and DNA likely occurred in solution as shown by AUC (Supplementary Figures 2, 3). Different species of protein–dsRNA or protein–DNA complexes were not distinctive by EMSA due to smearing (Figure 1A). The mechanism for association of more than one HBD with a 12 bp duplex will become obvious in the light of the crystal structure.

Structure determination of Hs-HBD complexed with a 12 bp RNA/DNA

Crystals of Hs-HBD complexed with the same 12 bp RNA/DNA as in the binding assays were readily formed and diffracted X-rays to 2.4 Å but with high mosaicity and smeared diffraction spots (Supplementary Figure 4A). Structure determination by molecular replacement using the apoprotein structure (PDB: 1QHK (Evans and Bycroft, 1999)) was unsuccessful. A 5-iodouridine was then introduced into the DNA strand for experimental phasing. This modification extended the resolution limit to 2.1 Å, but isomorphous replacement failed. The structure was eventually solved using the SeMet SAD method after introducing a Met (V71M) into the protein and crystallizing it in complex with a 6 bp RNA/DNA (Table I). This 6 bp complex structure could not be refined (see Supplementary data), but this initial model was then used in molecular replacement to solve and refine the structure of HBD complexed with the 12 bp RNA/DNA containing a 5-iodouridine (Table II). A sample of electron density map is shown in Supplementary Figure 4B.

Structure of the Hs-HBD–RNA/DNA complex and monomeric nature of HBD

Three Hs-HBDs are bound to each 12 bp RNA/DNA hybrid, and two such complexes are stacked head-to-tail in each asymmetric unit (Figure 2A). Each protein molecule directly interacts with 5 bp but spans 9 bp of the duplex along its length (~30 Å). The six protein molecules have little contact with each other and are virtually identical within the secondary structure elements (Supplementary Figure 5A). Large deviations are observed at the N- and C-termini of the domain and in the RNA-binding loop (see discussion later). The two

Table I SAD data collection and phasing statistics

	6-mer RNA/DNA complex with SeMet Hs-HBD
Space group	$p3_221$
Cell dimensions	
a, *b*, *c* (Å)	68.16, 68.16, 68.20
α, β, γ (deg)	90, 90, 120
Resolution (Å)[a]	20–2.7 (2.8–2.7)
R_{merge}[a]	5.8 (18.5)
$I/\sigma I$[a]	50.3 (7.3)
Completeness (%)[a]	97.9 (85.2)
Redundancy[a]	10.2 (7.7)
Solutions from SOLVE:	
Figure of merit = 0.31	
Solutions of 2 Se atoms with height/σ:14.9 and 13.7	

[a]Values in parentheses are of the highest resolution shell.

Table II Data collection and refinement statistics

	5-Iodouridine-modified 12-mer RNA/DNA complex of Hs-HBD
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	45.49, 64.26, 140.32
Resolution (Å)[a]	60–2.1 (2.2–2.1)
R_{merge}[a]	6.1 (39.6)
$I/\sigma I$[a]	13.8 (3.9)
Completeness (%)[a]	96.3 (95.6)
Redundancy[a]	3.6 (3.6)
Refinement	
Resolution (Å)	2.1
No. of reflections	24814
R_{work}/R_{free}	22.0/29.0
No. of atoms	3749
Protein	2462
Nucleic acids	1033
Water	252
B-factors	
Protein	34.6
Nucleic acids	26.1
Water	39.5
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (deg)	1.3
Ramachandran plot	91.5% most favoured 8.5% additionally allowed

[a]Values in parentheses are of the highest-resolution shell.

protein–RNA/DNA complexes are nearly superimposable except for the positions of the third HBD molecules (IIIA and IIIB) near the end of the duplexes. Molecules III also have poor electron density and high B-factors.

The structure of Hs-HBD consists of a three-stranded antiparallel β-sheet and two short helices (αA and αB) in the order of ββαβα (Figure 2A, B). Each eight-residue helix flanking the β-sheet is stabilized by a pair of aromatic side-chain interactions and capped at the C-terminus by an Arg residue (Figure 2C). The topology of Hs-HBD resembles the previously reported NMR structure of *Saccharomyces cerevisiae* HBD (Evans and Bycroft, 1999), but the two have significant differences in the secondary structures and loops

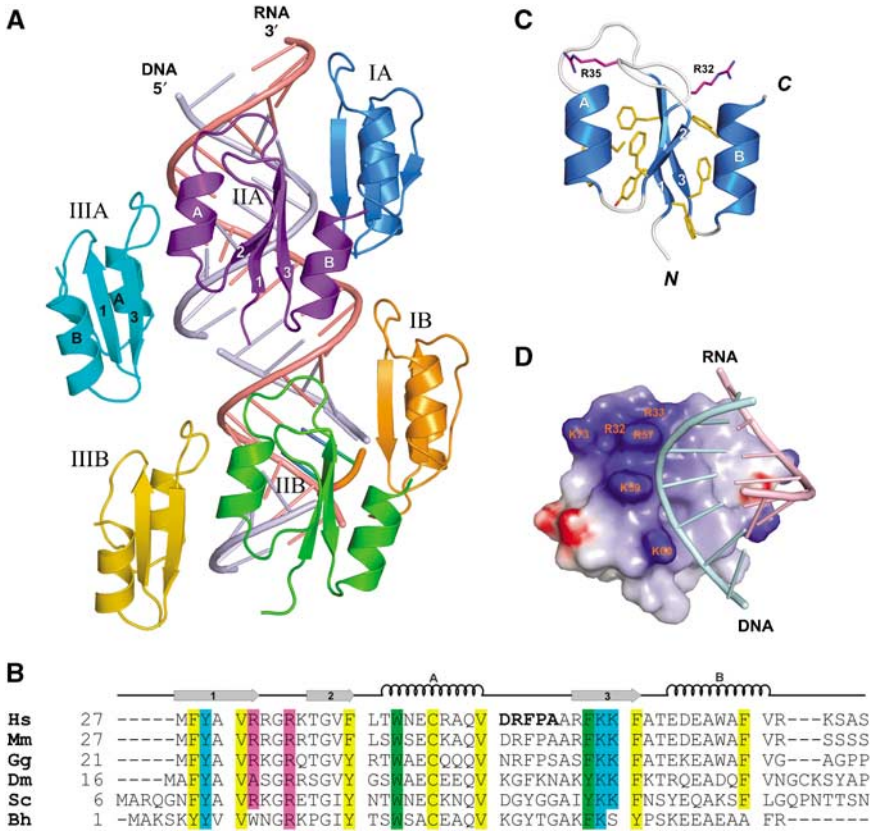

Figure 2 The structure of Hs-HBD complexed with a 12 bp RNA/DNA. (**A**) The content of an asymmetric unit. Each RNA/DNA hybrid (RNA in red, DNA in blue) interacts with three protein molecules. (**B**) Sequence alignment of HBD domains of RNases H1 from human (Hs), mouse (Mm), chicken (Gg), fruitfly (Dm), yeast (Sc) and *B. halodurans* (Bh). Conserved residues are highlighted in blue for DNA phosphate interaction, green for deoxyribose interaction, yellow for the hydrophobic core and purple for Arg's that stabilize the helices and RNA-binding loop. The RNA-binding loop is shown in boldface. (**C**) The structure of Hs-HBD. Side chains forming the hydrophobic core (mainly aromatic) and two Arg's capping the C-termini of helices are shown in sticks. (**D**) Electrostatic surface potential of Hs-HBD calculated using GRASP (Nicholls *et al*, 1991). The positively charged residues (blue) in a cluster are labelled.

(Supplementary Figure 5B). The RMSD over 40 pairs of Cα atoms between the crystal and NMR structures is 2.2 Å, which may explain the failure of molecular replacement. These differences may be due to the presence of RNA/DNA in the crystal structure or insufficient distance restraints for the NMR structure calculation.

The two RNA/DNA hybrids are pure A form with a narrow and deep major groove and shallow and wide minor groove and can be superimposed with the ideal dsRNA (Supplementary Figure 6). The backbones in the part of the hybrid that interact with protein molecules I and II are flattened leading to deformation of the double helix (Supplementary Figure 7). The deformation is not so obvious where molecule III binds near the end of the hybrid.

Accommodation of two and up to three HBDs on a 12 bp hybrid in the crystal structure is in agreement with the EMSA results (Figure 1A). The dimeric appearance of RNase H1 in the SPR experiments (Gaidamakov *et al*, 2005) is most probably due not to protein–protein interaction, as there is none, but to multiple loading of HBDs. Both HBD and RNase HC are monomeric in solution, and no positive cooperation is detected in HBD binding to RNA/DNA hybrid (Figure 1). In fact, concentration-dependence analyses of catalytic activity confirm that mammalian RNases H1 exist and function as monomers (Gaidamakov *et al*, unpublished data).

Recognition of an RNA/DNA hybrid

Hs-HBD interacts with the RNA/DNA hybrid mainly through αA and β3 (Figure 3A). All the protein–RNA/DNA interactions are with the backbones of the hybrid along the minor groove, as expected for non-sequence-specific recognition (Figure 2A, D). The protein and RNA/DNA contact in two regions—the loop between αA and β3 interacting with two consecutive 2′-OH groups of the RNA strand (Figure 3B), and a positively charged shallow groove interacting with three phosphodeoxyribose units of the DNA strand (Figure 3C). The two contact regions cover 5 bp. The RNA-binding loop consisting of residues D51–A56 (boldface in Figure 2B) is bolstered by the highly conserved R35, which forms a hydrogen bond with the carbonyl oxygen of D51. The main-chain amides of R52 and A55 form hydrogen bonds with the 2′-OH groups (Figure 3B). Neither phosphate groups of the RNA strand nor any protein side chains are involved in direct contact. This may explain why the structure but not the sequence of the RNA-binding loop is conserved (Figure 2B).

The DNA-binding surface is composed of both charge–charge interactions with the phosphate groups and aromatic stacking with deoxyriboses. The hydroxyl of a highly conserved Y29 (β1), the amine and amide of K60 (β3) and the guanidinium of R57 interact with three consecutive phos-

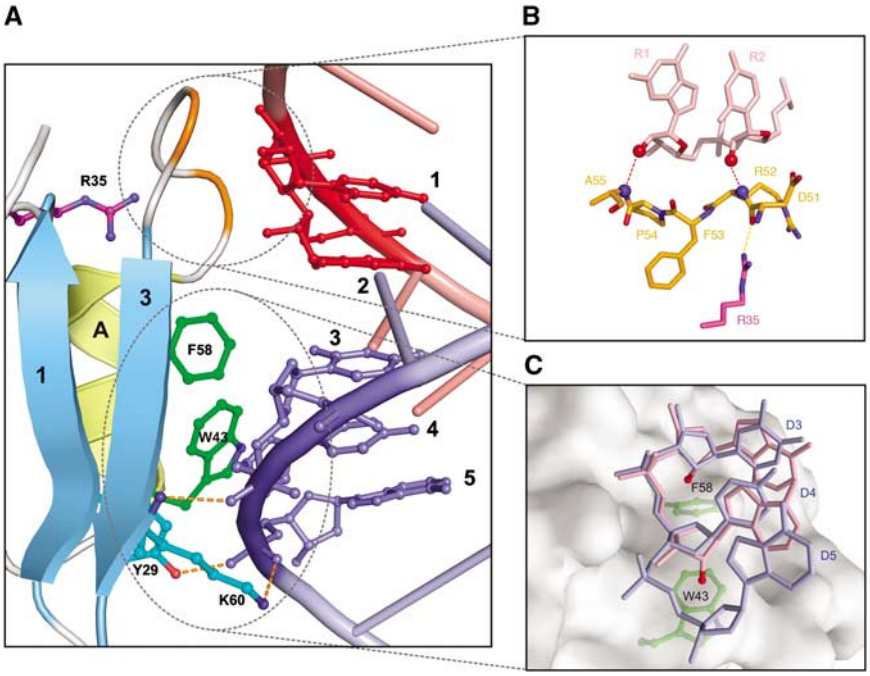

Figure 3 Interactions between HBD and RNA/DNA. (**A**) The nucleic-acid interface. Residues interacting with the DNA strand and R35 stabilizing the RNA-binding loop are shown in ball-and-sticks with the same colour coding as in Figure 2B. Base pairs interacting with HBD are numbered 1–5. (**B**) The close-up view of the RNA-binding loop. The 2′-OH groups interacting with the loop are shown as red spheres. The dashed lines indicate hydrogen bonds. (**C**) W43 and F58 shown in green ball-and-sticks under the molecular surface closely interact with deoxyriboses of the DNA (blue) and would clash with the 2′-OH groups (red) of a modelled RNA (light pink). Ribonucleotides (R1 and R2) and deoxyribonucleotides (D3-D5) are numbered as in (A).

phates of DNA. The aromatic rings of W43 (αA) and F58 (β3) form an aromatic patch and closely approach two consecutive deoxyribose rings of the DNA (Figure 3C). When an RNA strand is modelled into the DNA-interacting groove of Hs-HBD, the 2′-OH groups clash with F58 and W43 (Figure 3C). Thus, the interface formed by the two aromatic residues appears complementary to 2′-deoxyriboses only and leads to the preferential binding of RNA/DNA hybrids. Similar recognition mechanism for the A-form DNA was observed in the basic protrusion region of the catalytic domain of human RNase H1 (Nowotny *et al*, 2007).

Superimposition of the six HBD molecules in each asymmetric unit reveals that the protein–DNA interface is constant in spite of varied side-chain conformations of R57, K59 and K60. The Cα positions of the RNA-binding loop vary by up to 1.5 Å, but the protein movement is compensated by slight alterations of the RNA strand and the protein–RNA interface remains practically identical (Supplementary Figure 8). The observed weak binding of dsRNA by HBD probably requires displacement of the backbone of the second RNA strand from the DNA interface, which may be facilitated by the flexible RNA-binding loop and side chains of R57, K59 and K60.

Mutagenesis of residues involved in nucleic acid binding

The nucleic acid interface in Hs-HBD constitutes only $\sim$220 Å^2, and the most positively charged and semi-conserved surface area of Hs-HBD is not directly involved in RNA/DNA binding (Figure 2D). To verify the authenticity of the crystallographically observed RNA/DNA interface in HBD, the following substitutions were introduced into Hs-HBD: Y29F, KK59AA (K59 and K60 double mutation), R57A and W43A on the observed nucleic-acid interface, and double mutations of RR32AA and RK72AA outside of the interface. The ability of mutant Hs-HBD proteins to bind RNA/DNA hybrids and dsRNA were qualitatively examined by EMSA. All the mutations abolished dsRNA binding (data not shown) probably due to the inherent weak interactions of even the WT with dsRNA. As shown in Figure 4A, Ala replacement of W43 or K59 and K60, which are involved in DNA recognition, led to severely decreased RNA/DNA binding. Similar results were obtained with mouse RNase H1 (Gaidamakov *et al*, 2005). Binding of Hs-HBD to RNA/DNA hybrids was also compromised by the Y29F mutation. Although R57 lies in the vicinity of the DNA backbone, its substitution with Ala had only a small effect on hybrid binding. The RK72AA double-mutant protein is slightly defective in hybrid binding, even though these two surface residues are $>$10 Å away from the RNA/DNA. Their role may be to create an overall positively charged environment and allow non-specific approaches to nucleic acids before formation of the specific complexes. Based on the mutagenesis, we conclude that the observed HBD-RNA/DNA interface is genuine.

HBD contributes to substrate loading and processivity of RNase H1

To determine the contribution of HBD to the substrate binding and enzymatic activity of human RNase H1, all the above-mentioned amino-acid substitutions were introduced into the full-length protein. In addition, R35, which supports the RNA-binding loop, was substituted by Ala. Only the full-length protein of R35A could be produced in a soluble form in *E. coli*. The isolated Hs-HBD was insoluble probably because

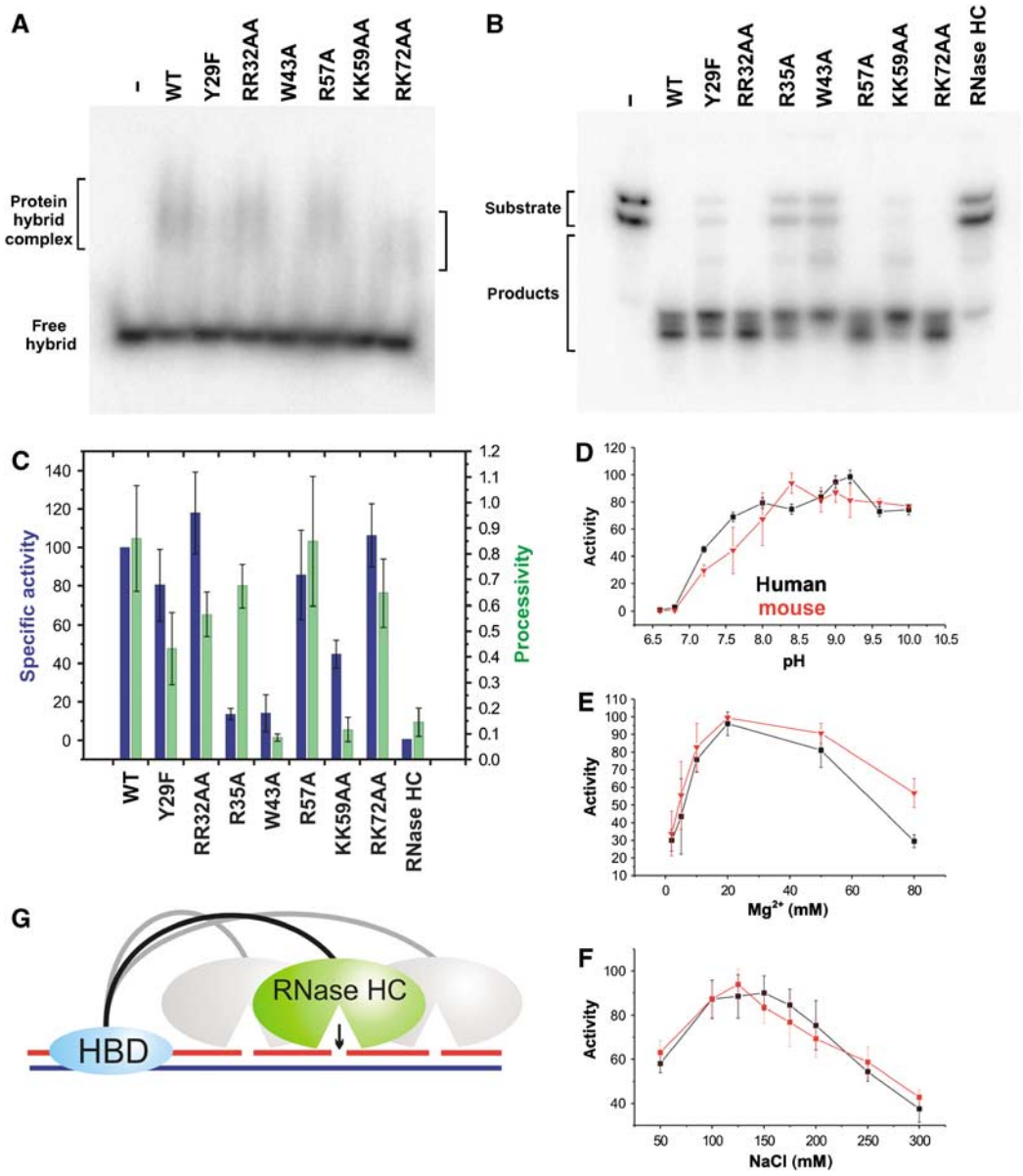

Figure 4 Mutagenesis studies and a model of RNA/DNA degradation by RNase H1. (**A**) EMSA of mutant Hs-HBDs. Mixtures of a mutant protein and the 12 bp hybrid (4 μM each) were resolved on a native TBE polyacrylamide gel. The smeared band corresponding to protein-hybrid complexes is indicated by brackets. (**B**) Cleavage assay of the 5′ end-labelled 36 bp RNA/DNA by RNase HC and full-length Hs-RNases H1 with mutations in HBD. The proteins are indicated on the top of the panel. The hybrid substrate has two bands. (**C**) A bar graph of specific activity (blue) and processivity (green) on the uniformly ^{32}P-labelled poly-rA/poly-dT substrate. The relative activity is normalized to WT RNase H1 (100%). The error bars in C to F represent the standard deviation of at least three independent measurements. (**D**) pH, (**E**) Mg^{2+} concentration and (**F**) salt concentration dependence of human (black) and mouse (red) RNase H1 activity on uniformly ^{32}P labelled poly-rA/poly-dT substrate. (**G**) A cartoon of RNA/DNA degradation by RNase H1. HBD (cyan) enhances substrate recognition and probably anchors the enzyme for multiple rounds of cleavage by RNase HC (green) leading to an increased processivity.

the mutation destabilizes the small domain, and RNase HC may stabilize the mutant HBD as a fusion partner. The hydrolytic activities of WT, mutant full-length proteins and RNase HC without HBD towards a 36 bp RNA/DNA hybrid of unique sequence were examined (Figure 4B). R35A and W43A RNases H1 as well as RNase HC exhibited a much lower specific activity (based on the disappearance of the 5′ ^{32}P-labelled 36 mer RNA) than WT and the remaining variant proteins.

The specific activities of mutant proteins were also examined using the uniformly ^{32}P-labelled poly-rA/poly-dT as substrate, and the results are in agreement with the cleavage results of the 36 bp hybrid (Figure 4C). Interestingly, the specific activity of WT human RNase H1 decreases dramatically in salt concentrations below physiological (Figure 4F), whereas RNase H1 with defective (W43A) or no HBD (RNase HC) loses catalytic activity with increased salt concentrations (Supplementary Figure 9). These observations support our speculation of the non-specific charge–charge interaction between HBD and nucleic acid, which appears to be dominant and interfering at low salt concentrations but enhances the catalytic activity of RNase H under physiological conditions.

Sensitivity to salt concentrations due to non-specific interactions also explains a previous mysterious observation that mammalian RNases H1 were most active in the presence of 50 mM Mg^{2+}, although the catalytic domain without HBD was inhibited (Gaidamakov *et al*, 2005). As the crystal structure of HBD provided no clue as to why non-physiologically high Mg^{2+} concentration was required, the optimal pH and Mg^{2+} in addition to the salt concentrations of mouse and human RNases H1 were systematically determined for the first time. The highest activity of both enzymes was observed at the physiological salt concentration (~125 mM NaCl), 10–20 mM $MgCl_2$ and pH above 7.9. Previously, the optimal $MgCl_2$ concentration of 50 mM was observed with 50 mM NaCl, lower than the optimal 125 mM (Gaidamakov *et al*, 2005). The high $MgCl_2$ concentration was likely needed to overcome the non-specific charge–charge interactions of HBD and nucleic acid (Figure 2D).

Cleavage assays of the 36 bp hybrid also revealed that W43A and KK59AA substitutions, which led to the most defective hybrid binding of HBD, caused a defect in generation of the smallest hydrolysis products (Figure 4B), suggesting impaired processivity. Processivity of the mutant proteins were examined in the following two experiments. Time points of degradation of the ^{32}P end-labelled 36 bp hybrids and uniformly ^{32}P-labelled poly-rA/poly-dT by varying concentrations of WT and mutant Hs-RNases H1 were taken as described previously (Gaidamakov *et al*, 2005). Processivity of RNase H was indicated by the ratio between the amounts of small- and intermediate-sized products after resolving hydrolysis products on a sequencing gel. In accordance with the hybrid-binding ability of HBDs, the RR32AA, R57A and RK72AA mutant proteins exhibited similar processivity to the WT, the Y29F RNase H1 had slightly reduced processivity, and the W43A and KK59AA proteins had as poor processivity as the isolated C-terminal domain (RNase HC) (Supplementary Figure 10, Figure 4C). The R35A RNase H1 had a low specific activity but near WT processivity, which may be attributed to the defective RNA-binding loop but intact DNA interface.

A model of RNA/DNA hybrid degradation by Hs-RNase H1

The processivity of mammalian RNase H1 was previously reported but ascribed as a result of dimerization through the HBD domains (Gaidamakov *et al*, 2005). As shown here, HBD is monomeric and increases the processivity of RNase H in RNA hydrolysis by its affinity for RNA/DNA hybrid. The analyses of RNA/DNA hydrolysis by WT and mutant RNases H1 suggest that HBD may have two important functions in facilitating the catalytic efficiency. Firstly, HBD enhances the affinity of RNase H1 for RNA/DNA hybrids. The K_d values of separate HBD (Figure 1B) and RNase HC (Nowotny *et al*, 2007) for RNA/DNA hybrid are ~100 nM, and by combining the two, the K_d of the full-length mouse RNase H1 is ~40 nM (Gaidamakov *et al*, 2005). Combination of multiple binding modules often leads to avidity and increased specificity, for example, multiple RNA-binding domains connected by flexible linkers in Staufen mRNA localization protein and ADAR2 RNA-editing enzyme (Lunde *et al*, 2007). Secondly, HBD enhances the processivity of RNase H1. Although the K_d values of HBD and a catalytic mutant RNase HC (Nowotny *et al*, 2007) for RNA/DNA hybrid are similar, nucleic acid binding by HBD is independent of catalysis and metal ions. When RNase HC is dissociated from the cleavage product after hydrolysis, HBD may anchor the enzyme, or at least facilitate its re-association, to the substrate to perform another round of cleavage in the vicinity of the first one (Figure 4G). Such 'processivity' enables WT protein to produce smaller-nucleotide products than the distributive catalytic domain alone.

Comparison of HBD with dsRNA and dsDNA binding modules

The mode of nucleic acid recognition by HBD is markedly different from protein domains that recognize dsRNA or dsDNA without sequence specificity. Non-sequence-specific binding of dsDNA, as observed in DNA polymerases, helicases, ligases and glycosylases, is often achieved by multiple helix–hairpin–helix (HhH) motifs, which bind B-form dsDNA across the minor groove (Doherty *et al*, 1996; Shao and Grishin, 2000). For example, BAF (barrier-to-autointegration factor) protein of ~90 residues contains two HhH motifs (Bradley *et al*, 2005). Each HhH motif approaches a DNA strand with the positive end of an α-helix dipole (Hol, 1985) and forms hydrogen bonds mainly between backbone amides and phosphosugar backbones (Figure 5A).

Previously, HBD had been compared with the dsRNA binding domain (dsRBD) of ~70 residues found in a variety of proteins, including related Dicer and RNase III nucleases (MacRae and Doudna, 2007), ADAR2, RNA-dependent protein kinase (PKR) and Staufen (St Johnston *et al*, 1992). HBD and dsRBD domains are similar in having a central three-stranded anti-parallel β-sheet and two helices, but the topology of dsRBD is different (αββββα), especially with the two helices on the same side of the β-sheet (Bycroft *et al*, 1995; Ramos *et al*, 2000; Ryter and Schultz, 1998) (Figure 5B and C). dsRBD specifically recognizes dsRNAs by contacting the phosphate groups with the N-termini of α-helices (as observed with BAF) and forming hydrogen bonds with the 2′-OH groups along the minor groove through protein side chains (Gan *et al*, 2006; Ryter and Schultz, 1998). Interestingly, the two α-helices of dsRBD interact with the two RNA backbones across the major groove, whose dimension is similar to the minor groove width of the B-form DNA (Figure 5C).

HBD is the only example of non-sequence-specific binding of nucleic acid without using an α-helix dipole. In addition to RNases H1, HBD is also found in cauliflower mosaic virus (CaMV) viroplasmin (also called TAV or ORF VI) (Mushegian *et al*, 1994; Cerritelli *et al*, 1998). CaMV is a dsDNA virus, but viral replication requires reverse transcription of the transcribed RNA. TAV is multifunctional and a major component of CaMV inclusion bodies, where replication, transcription and translation occur. The exact functions of HBD in TAV is not known, but its deletion prevents viral replication (Kobayashi and Hohn, 2003). Both the charged and aromatic residues in the DNA interface and the structure of the RNA-binding loop are conserved in HBD of CaMV. We anticipate that it specifically binds RNA/DNA hybrid. Given the dimeric nature of TAV, HBD may have an important function in reverse transcription and compaction of replication intermediates in the viral inclusion bodies.

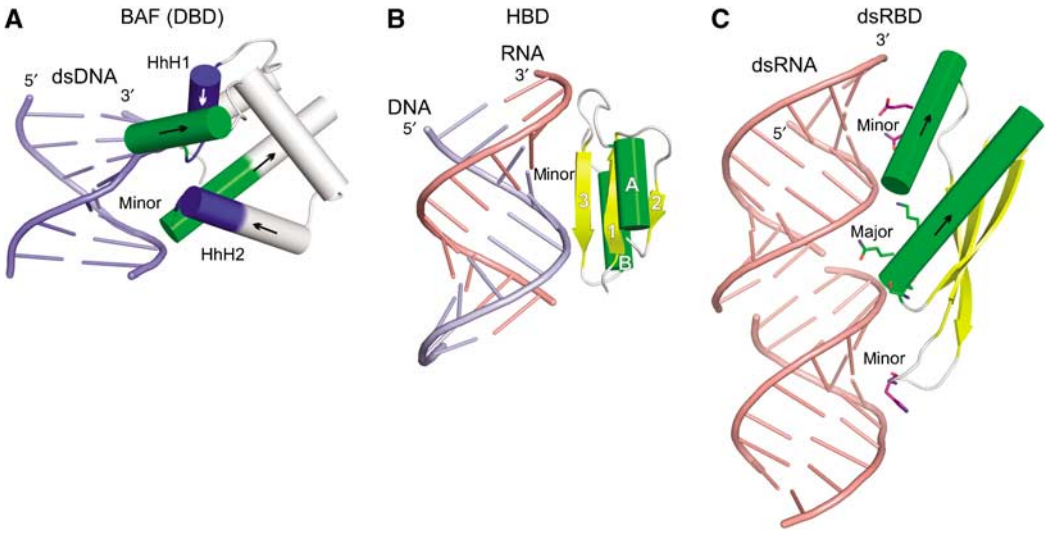

Figure 5 Non-sequence-specific binding of DNA, RNA/DNA and dsRNA. (**A**) BAF (PDB: 2BZF) with two HhH motifs (each with the first helix shown in blue and the second in green) interacting with DNA backbones across the minor groove. Most interactions are formed by the main-chain amides. (**B**) Hs-HBD reported in this study. (**C**) dsRBD from *Xenopus leavis* RNA-binding protein (PDB: 1DI2). It interacts with two successive minor grooves of the dsRNA, and two protein helical dipoles interact with the RNA backbones across the major groove.

Conclusions

It was remarkably difficult to obtain the crystal structure of HBD complexed with an RNA/DNA hybrid. In fact, small and flexible proteins as HBD are notoriously difficult to crystallize, and often the way to overcome this problem is to induce oligomerization. Previous examples include hexameric insulin (Blundell *et al*, 1971) and multimeric zinc fingers (Pavletich and Pabo, 1991). Our experience with HBD suggests that the double helix can be a scaffold for oligomerization of small nucleic acid-binding proteins and enhance the crystallization probability.

Our structural and biochemical analyses of HBD also show that a small substrate-binding domain tethered to a catalytic domain can enhance both the specific activity and processivity of the enzyme. It is predicted that the shorter the linker between two binding modules, the higher the avidity (Shamoo *et al*, 1995). HBD linked to RNase HC by ~10 residues in *B. halodurans* RNase H and dsRBDs closely coupled to the catalytic domains in RNase III and Dicer may greatly increase the substrate binding of these enzymes. The long linkers (60 aa) in human and mouse RNases H1 appear to enhance substrate binding rather moderately (~2.5-fold), but a long linker, which allows the catalytic domain to access the substrate at multiple sites, may be required for the increased processivity. Long linkers between the dsRBDs and catalytic domain in ADAR2 suggest that adenosine deamination of RNA by ADAR2 is processive.

Materials and methods

DNA constructs and protein purification

The region encoding residues 27–76 of human RNase H1 was amplified by PCR from a full-length cDNA (Cerritelli and Crouch, 1998) and cloned into pET15b. Mutations were generated using Quikchange (Stratagene). The N-terminal His_6-tag HBD proteins were expressed in *E. coli* BL21 (DE3) Rosetta cells (EMD Chemicals) in LB medium by induction with 0.4 mM IPTG after growth at 37°C reached 0.7 OD_{600} and purified on nickel and Mono S columns (see Supplementary data for details).

Crystallization

RNA oligonucleotides were purchased from Dharmacon and DNA oligos from IDT (Supplementary Table I). Oligo RNAs were deprotected according to the manufacturer's protocol and desalted (6 mer RNA) or purified using urea-PAGE (12 mer). The RNAs were eluted from gel fragments in the buffer containing 10 mM Tris (pH 8.0), 300 mM NaCl and 5 mM EDTA. Oligo DNAs (with trityl group on) were HPLC-purified using a reverse-phase R3 column (Applied Biosystems) and detritylated. Complementary RNAs and DNAs were annealed by heating to 65°C and slow cooling.

Native or SeMet Hs-HBD (4–5 mg/ml final concentration) were mixed with the RNA/DNA hybrids at 1:1 molar ratio for the 6 bp RNA/DNA and 2:1 molar ratio for the 12 bp RNA/DNA. The complexes were mixed with the reservoir solution at equal volume and crystallized by the hanging drop vapour diffusion method at 21°C. The 6 bp complex crystals were obtained with the well solution containing 10% PEG3350, 0.2 M NaCl, 0.1 M Tris (pH 8.5). Crystals were cryoprotected by a stepwise increase of sucrose concentration up to 25% and flash-frozen in liquid nitrogen. The 12 bp complex crystals were obtained with 1.2 M NaCl and 0.1 M HEPES (pH 7.5). The crystals were transferred to Paratone oil for cryoprotection and flash-frozen before diffraction.

Data collection and structure determination

The X-ray diffraction data were collected at ID-22 beamline at Advanced Photon Source on a Mar225 or Mar300 CCD detector at 100 K. The wavelength was 0.97928 Å for the SAD experiment and 1 Å otherwise. The 12 bp complex data were processed and scaled with XDS (Kabsch, 1993) (Table I). The structures of Hs-HBD in complex with RNA/DNA were traced in COOT (Emsley and Cowtan, 2004) and O (Jones *et al*, 1991) and the resulting models were refined using CNS (Brunger *et al*, 1998) interspersed with manual building (for details, see Supplementary data).

Electrophoretic mobility shift assays

WT and mutant Hs-HBD proteins were mixed with the cold 12 bp RNA/DNA hybrid, dsRNA or dsDNA at a 1:1 molar ratio. Serial twofold dilutions of the complex (ranging from 32.0 to 0.031 μM) were prepared in the reaction buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 5% glycerol, 10 mM DTT and 0.5 mM EDTA). For visualization, ^{32}P-labelled 12 bp duplex was added to each reaction at a final concentration of ~5 nM. The samples were resolved on a 6% polyacrylamide gel in 0.5 × TBE buffer at 4°C and visualized by phosphoimaging. Essentially, the same results were obtained when radioactive and cold nucleic acid were mixed first and the protein added last, indicating a fast exchange in the complexes.

Sedimentation equilibrium

Detailed description of sedimentation equilibrium experiments can be found in Supplementary data. Briefly, the experiments were conducted at 4°C on a Beckman Optima XL-A analytical ultracentrifuge at rotor speeds of 16–28 k.r.p.m. Samples containing 1:1 and 3:1 stoichiometric mixtures of the Hs-HBD and the 12 bp RNA/DNA, dsRNA or dsDNA (loading volume of 130 μl) were studied at ~1.7, 3.4 and 5.1 μM. Data were acquired using six-hole cells at an average of four absorbance measurements using a radial spacing of 0.001 cm and analyzed globally in terms of various species analysis models using SEDPHAT 4.3 (www.analyticalultracentrifugation.com/sedphat/). Interaction affinities were determined using $A+B \leftrightarrows AB$ and $A+B+B \leftrightarrows AB+B \leftrightarrows ABB$ reversible equilibrium models. To account for the fivefold differences in the extinction coefficients at 280 nm, the protein-to-nucleic acid ratios were fixed to 1 or 3 in the implementation of the mass conservation constraint.

Specific RNase H activity

WT and mutant proteins were assayed using either a 36 bp hybrid with the RNA strand 5′ ^{32}P-labelled (see Supplementary Table I for sequences) or poly-rA/poly-dT with uniformly labelled poly-rA. Aliquots of the 36 bp hydrolysis reaction, which contained 2.8 nM substrate and 0.28 nM each protein in 50 mM Tris–HCl (pH 7.9), 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 20 μg/ml BSA and 5% glycerol and were incubated at 22°C for 15 min, were applied to a 20% TBE-urea polyacrylamide gel. The reaction products were visualized by phosphoimaging. With the poly-rA/poly-dT substrate, assays were performed in a total volume of 10 μl containing 0.5 μM ^{32}P-labelled poly-rA/poly-dT in 50 mM Tris–HCl, pH 7.9, 150 mM NaCl, 1 mM DTT and 5 mM $MgCl_2$. Reactions were initiated by adding WT or mutant Hs-RNase H1 to a final concentration of 4 pM, or by adding RNase HC (8 pM), and terminated after 30 min incubation at 37°C by chilling on ice and followed by additions of 10 μl of 10 mg/ml BSA and 100 μl of 20% TCA. After 15 min on ice, tubes were centrifuged at 16 000 *g* and 100 μl of supernatant containing reaction products less than 20-nucleotides long was counted in liquid scintillation counter.

Processivity assays with poly-rA/poly-dT and 36 bp hybrid of unique sequence

The assays were conducted as described previously (Gaidamakov *et al*, 2005). Briefly, enzyme in a range between 0.1 and 1 nM for wild-type and HBD mutant RNase H1 and 8 nM for RNase HC (amounts of the enzymes were pre-calibrated to achieve comparable level of substrate degradation) were mixed with 3 μM ^{32}P-labelled poly-rA/poly-dT or ~10 nM gel-purified 36 bp RNA/DNA hybrid in 50 mM Tris–HCl, pH 7.9, 50 mM NaCl, 1 mM DTT, 20 μg/ml BSA, 4% glycerol. Reactions were initiated by addition of 5 mM $MgCl_2$. With poly-rA/poly-dT reactions, samples of 10 μl were removed at time periods ranging from 7 to 240 s and cleavage products were resolved on sequencing gels. Final products were defined as oligonucleotides 12 nucleotides or shorter in length, and intermediate products ranged from 13 to ~60 nt in lanes where less than 85% of the starting substrate was degraded. Processivity was computed by dividing the amounts of final to intermediate products. For mutant and WT Hs-RNases H1, each number shown in Figure 4C was an average of 4–8 data points.

Salt, Mg^{2+} and pH optimization

Specific activities of human and mouse RNases H1 were assayed with ^{32}P-labelled poly-rA/poly-dT as described above. To determine optimal salt concentrations initially, enzyme at 4 pM final concentration was added to 10 μl solution containing 0.5 μM substrate in 50 mM Tris–HCl, pH 7.9, 1 mM DTT, 5 mM $MgCl_2$ and 50, 100, 150 or 200 mM NaCl. The optimal NaCl was found to be 150 mM. To determine the pH optima, the same reactions were conducted in 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 50 mM buffer solution with pH ranging from 6.6 to 10.0, using sodium cacodylate trihydrate for pH 6.6 and 6.8, HEPES sodium for pH 7.2 and 7.6, Tris–HCl for pH 8.0, 8.4 and 8.8, CAPSO for pH 9.0, 9.2 and 9.6 and CAPS for pH 10.0. A pH value of 7.9 and above appeared to be optimal. To determine the optimal Mg^{2+} concentration, assays were performed in 50 mM Tris–HCl, pH 7.9, 150 mM NaCl and 1 mM DTT with 2–80 mM $MgCl_2$. The Mg^{2+} concentrations of 10–20 mM were optimal. The final NaCl optimization was determined in 50 mM Tris–HCl, pH 7.9, 10 mM $MgCl_2$, 1 mM DTT and NaCl ranging from 50 to 300 mM.

Accession codes

Protein Data Bank: Coordinates and structure factors have been deposited with accession code 3BSU.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Lisa Eiben for excellent technical assistance and the staff of SERCAT beamline at Advanced Photon Source for assistance with data collection. This research was supported by the Intramural Research Program of the NIH, NIDDK, NICHD and Intramural AIDS Targeted Antiretroviral Program.

References

Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DC, Mercola DA, Vijayan M (1971) Atomic positions in rhombohedral 2-zinc insulin crystals. *Nature* **231:** 506–511

Bradley CM, Ronning DR, Ghirlando R, Craigie R, Dyda F (2005) Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat Struct Mol Biol* **12:** 935–936

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54** (Part 5)**:** 905–921

Bycroft M, Grunert S, Murzin AG, Proctor M, St Johnston D (1995) NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J* **14:** 3563–3571

Cerritelli SM, Crouch RJ (1995) The non-RNase H domain of *Saccharomyces cerevisiae* RNase H1 binds double-stranded RNA: magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA* **1:** 246–259

Cerritelli SM, Crouch RJ (1998) Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase HI. *Genomics* **53:** 300–307

Cerritelli SM, Fedoroff OY, Reid BR, Crouch RJ (1998) A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucleic Acids Res* **26:** 1834–1840

Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, Crouch RJ (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell* **11:** 807–815

Crouch RJ, Cerritelli SM (1998) Ribonucleases H of *Saccharomyces cerevisiar*, *Schizosaccharomyces pombe*, *Crithidia fasciculata* and *Neurospora crassa*. In Crouch RJ, Toulme JJ (eds.), *Ribonucleases H*. Paris: INSERM, pp 79–100

Doherty AJ, Serpell LC, Ponting CP (1996) The helix–hairpin–helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res* **24:** 2488–2497

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Evans SP, Bycroft M (1999) NMR structure of the N-terminal domain of *Saccharomyces cerevisiae* RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9. *J Mol Biol* **291:** 661–669

Gaidamakov SA, Gorshkova II, Schuck P, Steinbach PJ, Yamada H, Crouch RJ, Cerritelli SM (2005) Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Res* **33:** 2166–2175

Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124:** 355–366

Hol WG (1985) Effects of the alpha-helix dipole upon the functioning and structure of proteins and peptides. *Adv Biophys* **19:** 133–165

Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47** (Part 2)**:** 110–119

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26:** 795–800

Kobayashi K, Hohn T (2003) Dissection of cauliflower mosaic virus transactivator/viroplasmin reveals distinct essential functions in basic virus replication. *J Virol* **77:** 8577–8583

Kogoma T, Foster PL (1998) Physiological functions of *E. coli* RNase HI. In Crouch RJ, Toulme JJ (eds.), *Ribonucleases H*. Paris: INSERM, pp 39–66

Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8:** 479–490

MacRae IJ, Doudna JA (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol* **17:** 138–145

Mushegian AR, Edskes HK, Koonin EV (1994) Eukaryotic RNAse H shares a conserved domain with caulimovirus proteins that facilitate translation of polycistronic RNA. *Nucleic Acids Res* **22:** 4163–4166

Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296

Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121:** 1005–1016

Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell* **28:** 264–276

Nowotny M, Yang W (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J* **25:** 1924–1933

Ohtani N, Haruki M, Morikawa M, Crouch RJ, Itaya M, Kanaya S (1999) Identification of the genes encoding Mn2 + -dependent RNase HII and Mg2 + -dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. *Biochemistry* **38:** 605–618

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252:** 809–817

Ramos A, Grunert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* **19:** 997–1009

Ryter JM, Schultz SC (1998) Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* **17:** 7505–7513

Shamoo Y, Abdul-Manan N, Williams KR (1995) Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res* **23:** 725–728

Shao X, Grishin NV (2000) Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res* **28:** 2643–2650

St Johnston D, Brown NH, Gall JG, Jantsch M (1992) A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci USA* **89:** 10979–10983